한국항행학회 논문지 제14권 제6호 2010년 12월

IMT Advanced 시스템에서 예측 스케줄링을 통한 핸드오버시 모바일 QoS 보존 방법

Preserving Mobile QoS during Handover via Predictive Scheduling in IMT Advanced System

니라지 포댤*, 이병섭**

Neeraj Poudyal*, Byung-Seub Lee**

요 약

본 논문에서의 새로운 스케줄링은 모바일 간에 실시간으로 요구하는 모든 최소 가용 대역폭, 최대 지연 시간 그리고 다른 부수적인 조건들에 대해 QoS를 보장하는 핸드오버 요청들을 제공하기 위해 제안되었다. 입력으로는 모바일 스테이션이 서비스지역 이동간에서 발생하는 핸드오버 시간에 대한 예측이 필요하다. QoS와 핸드오버 시간을 알게 된 후, 목적지 기지국은 새로운 요청을 위한 핸드오버에 우선순위를 둔다. 그리고 이전의 리소스는 실제로 핸드오버 동안 사용할 수 있는 기회를 조금 더 가지게 된다.

Abstract

In this paper, a novel schedulability criteria is developed to provide handover calls with Quality of Service (QoS) guarantees in terms of both minimum available bandwidth, maximum tolerated packet delay, and other additive QoS constraints as required by the real-time mobile traffic. This requires prediction of the handover time using mobility trends on the mobile station, which is used as input to this work. After the handover time and the QoS are negotiated, the destination base station makes attempts to give priority to handover calls over new calls, and pre-reserves resources that will have more chance of being available during the actual handover.

Key words : IEEE 802.16m, QoS Handover, Handover Prediction, Admission Control, IMT-Advanced

I. Introduction

The purpose of this paper is to provide deterministic Quality of Service (QoS) guarantees to the mobile real time flows during handover when some QoS requirements and handover time are predicted in advance. We choose IEEE 802.16m [1] as our preferred IMT Advanced [2,3] technology. In order to maintain the QoS during the handover, the system should have enough resources, and also the switchover should be fast so that the handover request's delay guarantees are maintained. Dynamic pre-reservation of resources for handover calls is necessary for such deterministic guarantees. Handover request queuing is good for

* 한국항공대학교대학원 정보통신공학과 (Information & Telecommunication Engineering, Graduate School, Korea Aerospace University)
** 한국항공대학교 정보통신공학과 (Information & Telecommunication Engineering, Korea Aerospace University)
· 제1저자 (First Author) : Neeraj Poudyal
· 투고일자 : 2010년 10월 27일
· 심사(수정)일자 : 2010년 10월 28일 (수정일자 : 2010년 12월 6일)
· 게재일자 : 2010년 12월 30일

enhancing handover performance for best effort traffic, but they slow down the handover time and are not suited for real time traffic. Newer papers such as [4] address location tracking methods such as using the trajectory of mobile nodes to predict the destination base station as well as time of handover using media independent handover services [5] prior to actual handover. In [4], the authors propose an information server where some information such as velocity, position, movement detection, etc. is recorded for mobile terminal having GPS capability to start the handover procedure. Other mobility prediction based handover approaches have been studied in [6-9]. Our paper does not re-address the specifics of such a prediction. It instead addresses the problem of providing QoS guarantees after the handover time has been predicted.

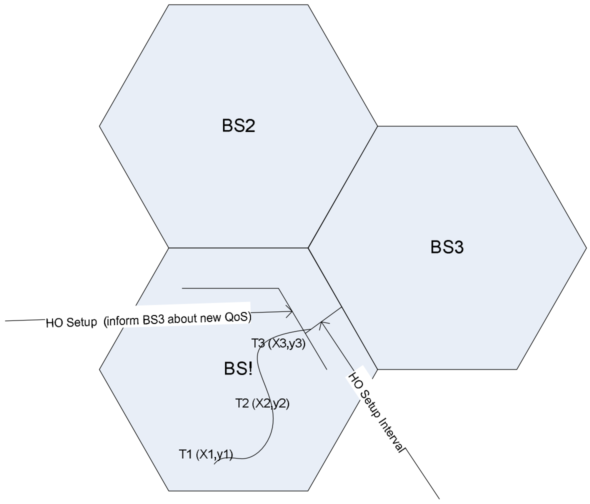

그림 1. 목적지 기지국으로의 예측 핸드오버시 궤적
Fig. 1. Trajectory used for prediction of destination base station and handover time in advance

Ⅱ. System Model

IEEE 802.16m refers to the base station and mobile station that support this standard as advance base station (ABS)[1] and advanced mobile station(AMS)[1] to differentiate itself from its older versions IEEE802.16a-e. In this paper we use the names interchangeably. Unless otherwise stated, destination base station will refer to the user's targeted ABS to which the connection is handed over from the source ABS, rather than the base station that receives the end traffic. For multi-hop QoS analysis, we assume every ABS is connected to the wired. Every time an advanced mobile station (AMS)[1] requests bandwidth, the admission control procedure of the advanced base station(ABS) checks whether its QoS parameters such as minimum sustained rate and maximum tolerated delay can be met before the traffic is admitted. This is true for both handover and new calls. We assume that there is a precise cellular boundary in the source base station region which is close to the destination base station (fig.1).

Depending on its direction, movement and speed statistics of the mobile trajectory, a nearly perfect prediction of the destination base station is performed. Handover prediction is however not the goal that is addressed in this paper. We simply assume the predicted data is available and it is used as input to our work. This advance knowledge is used by the destination base station to set up the handover procedure if the system can meet the QoS requirements. The system performs suspension scheduling where predicted handover requests are suspended and pre-reservation of resources are periodically attempted before the actual handover takes place. Whenever there are handover requests in the suspension list, they are given more priority than the new calls. The algorithm however does not affect the minimum sustained rate of the new calls, as they must be met as negotiated between the operator and the user through the Service License Agreement (SLA).

Ⅲ. QoS ANALYSIS

Different traffic flows have different QoS requirements. For example ftp traffic cannot tolerate loss, but it can tolerate some packet delay. Real time traffic on the other hand needs stringent delay bounds, although it can tolerate some packet loss. eg:, for a strict quality of service of a real time flow, packet delay must have deterministic guarantees whereas packet loss can have probabilistic guarantees. In this paper we derive such a bound for delay. Traditionally mobile systems have been used to serve voice traffic, and the way to provide QoS to these traffic is by calculating blocking probability of the base station by using (2) and then designing the system to have a very low blocking probability, for example less than .005. Assuming Poisson arrivals, arrival rate λ and departure rate μ, the probability that there are k calls in the system with traffic intensity a=λ/μ is given by the following equation:

$$P_k = (a^k/k!)/(\sum_{k=0}^{n} a^k/k!) \qquad (1)$$

The blocking probability B of the station is simply the probability that an arriving call n=k+1 will find all k channels busy. From Erlang B formula, the blocking probability is

$$B(n,a) = (a^n/n!)/(\sum_{k=0}^{n} a^k/k!) \qquad (2)$$

The use of (2) is however not efficient for data traffic because of its characteristic burstiness. Also such probabilistic measures cannot guarantee delay and other QoS parameters for the real time traffic. Providing QoS guarantees is done through network admission control [10-13], however, most of them provide only bandwidth guarantees and lack multiple additive QoS guarantees. IMT advanced system also requires various operation level QoS for the real time flows such as minimum sustained rate, maximum user rate and maximum tolerated packet delay. In this paper, we address these parameters for new as well as predicted handover calls. For a single base station analysis the following terms are used.

Terminology

Br=Bandwidth request. Br is removed after it is served

Br_k=Bandwidth request of already admitted connection k

Br_i=Bandwidth request of the new connection i

B_{total}=Total IEEE802.16m system bandwidth (measured) without considering network links

B_{usd}=Bandwidth occupied(measured),which is the sum of all Br_k

B_{avail}=B_{total}-B_{usd}

D_i=Maximum Delay requested by the new request i

C=Total number of pending requests, each request is removed from C when it is served

N=Number of connections

$aik(u)$ and $aik(d)$=k^{th} additive constraint for uplink and downlink

R_k=the rate of the total 802.16m system measured with the modulation rate of the station k. For example if the measured rate(capacity) of the IEEE802.16m system is 10Mbps at 64QAM(3/4) and 8Mbps at 64QAM(1/2), and the modulation rate of station k is 64QAM(1/2) as dictated by its SNR, then R_k=8Mbps.

Γ_i = Maximum user rate for connection i

γ_j= minimum sustained rate of connection j

In addition, for multi-hop analysis the following terms are used.

Bp=Bandwidth of a link p in P

P=Network path having Hhops,

p=Links in P including network links as well as the IEEE802.16m wireless uplink and downlink

$aik(p)$ and $Aik(P)$=k^{th} additive constraint for link p, and for path P

$\Pr(lip)$ and Li=packet loss probability for link p, and for path P

$Min(Bp)$=Lowest bandwidth among all the p in P

The QoS requirement during handover is usually sent as rate(bandwidth) and one or more additive constraints. In real time polling service class (rtPS), the additive constraint mainly consists of delay.

Rate of IEEE 802.16m: The basic unit of transmission in the physical layer is known as a symbol which can carry variable number of bits depending on various factors. The time to transmit each OFDM symbol with sampling frequency fs , fast Fourier transform at non equispaced nodes (NFFT) and cyclic prefix ratio G can be used to derive the rate of the system.

$$Symbol\ Time = 1/(fs/NFFT) \times (1+G) \quad (3)$$

The rate of transmission between the ABS and AMS is chosen by the Adaptive Modulation and Coding (AMC) Scheme depending on the received Signal to Noise Ratio (SNR) of the AMS. Depending on the Modulation and DVB-H coding, the useful bits per symbol (ubps) varies as shown in the example table 1.

표 1. 적응 변조 및 코딩 방식
Table 1. Adaptive Modulation and Coding Scheme

SNR	Modulation	Coding	Useful Bits Per Symbols (ubps)
6.0	QPSK	1/2	192*2*1/2=192
8.5	QPSK	3/4	192*2*3/4=288
11.5	16QAM	1/2	192*4*1/2=384
15.0	16QAM	3/4	192*4*3/4=576
19.0	64QAM	2/3	192*6*2/3=768
21.0	64QAM	3/4	192*6*3/4=864

$$Rate = Nmber\,of\,symbols \times ubps \quad (4)$$
$$Rate = \frac{1}{1/(fs/NFFT) \times (1+G)} \times ubps$$

(4) provides a theoretical system rate. In our implementation, we use measured rate rather than theoretical rate for accuracy. As can be seen in (4), the rate of the IEEE system is variable depending on the signal to noise ratio of the received signals from the AMSes. A fraction of this rate is reserved by the system for real-time traffic, which is classified under IEEE 802.16m system as real time polling services (rtPS) class [1]. Since this is the total rate available for the rtPS traffic, it is denoted by B_{total}. As discussed, the rate at which this transmission takes place depends on the robustness of the modulation scheme, defined by the AMC. Note that most previous papers [10], [11], [12], [13] do not consider AMC into account whereas our algorithm measures the data rate considering the effect of AMC implicitly. Inamulti-hop network, rate also depends on the capacity of each link p in path P.

Delay: End to end delay is a combination of optional channel acquisition delay, a polling delay of $n^* T_f$ where n is defined by the operator, queuing delays at the subscriber station and base station, and MAC-layer transmission delay. We ignore packet preparation and packet processing delay in our analysis assuming them to be negligible and not within the focus area of the network and MAC layers. Transmission delay depends on the rate of the total WiMax as defined by the AMC. In QoS systems, the operator usually sets the maximum user rate for fairness purpose.

Packet loss: Packet loss is not a strict requirement of rtPS traffic class, thus a probabilistic measure can be used for packet loss. The probability of packet loss at hop k multiplied by the bandwidth gives the number of packets lost. Packet loss is a multiplicative measure. Multiplicative additive constraints such as probability of packet loss can be converted to additive constraints by using logarithmic manipulation (as shown in (5) within in the curly braces) which is much easier for calculation. This paper does not explore packet loss further and just considers them to be one of the other additive constraints.

$$Li = \prod_{p=1}^{P} (1 - \Pr(li_p)) \times Br_i \quad (5)$$
$$i.e, Li = -(\sum_{p=1}^{P} \log(1 - \Pr(li_p)) \times Br_i$$

Other additive constraints: Since 802.16m requires guaranteeing bandwidth and delay only, this paper doesn't concern with other individual constraints. But they are still acknowledged and denoted as $Aij(P)$ where connection i is requesting j^{th} additive constraint in path $P(aij(p)$ refers to j^{th} constraint in link p for connection i). Define an operator Φ such that $A\Phi B$ returns true if both A and B are true, false otherwise. The additive constraints requirement of P can be written as:

$$\Phi_{k=1}^{MAXCONSTRAINTS} Aik(P)^{+} \geq \sum_{p=1}^{P} aik(p) \qquad (6)$$

It should be noted that delay is also a kind of additive constraint. Other examples of additive constraints are hop-count or packet loss. However, it is important to deal with delay as a separate additive constraint because unlike the other additive constraints, delay of a new request depends directly on the previously admitted bandwidth requests. In other words delay is not memory-less constraint like hop count and packet loss probability. Since IEEE 802.16m requires guaranteeing bandwidth and delay only, other constraints are not dealt with individually.

Ⅳ. ADMISSION CONTROL CRITERIA

We now formalize the admission control procedure. Given a path P from the source 802.16m AMS to the destination 802.16m AMS in an integrated network, the system needs to ascertain that P meets the following sets of QoS constraints:

1. Minimum sustained rate constraint of the k^{th} AMS as defined by the AMS's SLA is maintained. In other words, the system must maintain a reserved rate ψ such that $\psi \geq \sum_{k=1}^{N} \gamma_k$ where γ_k the minimum sustained rate of k^{th} AMS, and N is the number of active AMSes. 2. Bandwidth request is large enough to be transmitted through the network without impacting the minimum sustained rate of other flows. At the same time, each AMS is can have a configurable maximum rate Γ_i so that no single AMS can take unfair advantage of system bandwidth. 3. Deadline of the bandwidth request is large enough to be met by the multi-hop network. 4. A number of unspecified additive constraints such as hop delay, packet loss etc. is supported. From the explanation and observations above, we derive a schedulability condition. The admission control criteria for both new and handover calls are the same, except for the fact that the new calls will see a smaller B_{total} when there are handover calls in the suspension set. The admission control procedure for a multi-hop system with a guaranteed minimum sustained rate, maximum user rate, delay and additive QoS constraints are:

$$\begin{aligned}
& Br_i \leq MIN(\Gamma_i, B_{total} - \sum_{k=1}^{C} Br_k - \sum_{j=1}^{N} \gamma_j, \\
& MIN(Bp \,\forall\, p \in P) - \sum_{k=1}^{C} Br_k - \sum_{\substack{j=1, \\ j \neq k}}^{N} \gamma_j) \;\Phi \\
& D_i \geq (n+1)\, T_f + \sum_{p=1}^{P} \frac{Br_i + \sum_{k=1}^{C} Br_k}{B_p} + \\
& Ceil\left(\frac{1}{T_f}\left(\sum_{k=1}^{C} \frac{Br_k}{R_k} + \frac{Br_i}{\Gamma_i}\right)\right) T_f \qquad (7) \\
& \Phi_{K=1}^{MAXCONSTRAINTS} (Aik(P)^{+} \geq \sum_{p=1}^{P} aik(p)) \\
& D_i \in N \times T_f,\, N = 3, 4, \ldots
\end{aligned}$$

The available bandwidth in (7) is the minimum of maximum user rate, available IEEE 802.16m system bandwidth and the minimum of path bandwidths after reserving the minimum sustained rate of connected AMSes. The delay inequality (7) includes total end to end delay including maximum polling delay of nT_f where n is defined by the operator and channel access delay of $1T_f$. The Ceil function then converts the delay into the whole number of frames.

Ⅴ. SYSTEM IMPLEMENTATION

IEEE 802.16m system does not define a specific admission control. Therefore we add our own admission control using the criteria (7) and scheduling as shown in the fig. The same QoS admission control procedure is used for both the new call and the handoff call. The difference is in the available system bandwidth seen by the handover and the new calls.

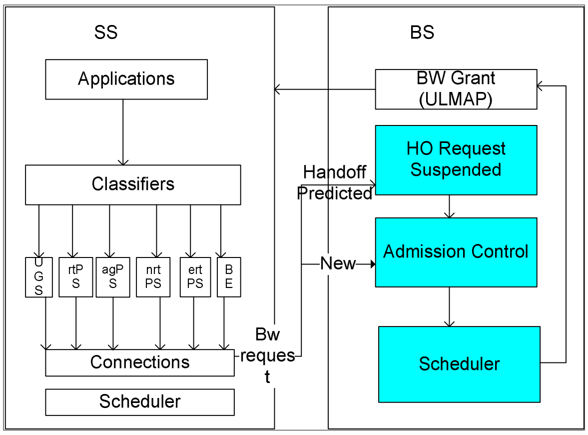

그림 2. 새로운 수락제어방식과 HO
Fig. 2. Admission Control for New and HO

Upon predicting the mobile node's destination base station, the source base station provides the destination with the QoS and bandwidth requirements and the relative predicted handoff time in number of frames. In other words there is an exchange of the QoS request set $R(B, QoS, T_h(p))$ between source and destination base station where B stands for requested bandwidth, QoS is requested QoS parameters such as delay, packet loss etc, and $T_h(p)$ is the predicted handover time in number of frames. The absolute predicted handoff time is $t+T_h(p)$ where t is the current time.

This information is queued in the destination base station until current time equals $t+T_h(p)-T_r(p)$ where $T_r(p)$ is the operator defined time in number of frames when the ABS starts to reserve channel for the handoff call. During full utilization, if required, new calls are not accepted to make room for the handoff calls. After the reservation timer begins, admission control at the destination base station is evoked every frame for $T_r(p)$ frames when the base station attempts to reserve the slot for the handover. If it is not successful at frame i, it will not accept a new call and reattempts at frame $i+1$. At the same time, some resources held by the existing calls will be continuously released increasing the likelihood of handover success every frame. If successful, when the actual handover request is made, the channel will have already been reserved for QoS handoff to take place. The algorithm used by scheduler is provided here. Note we use only delay and not other QoS metrics in simulation. Also the handoff calls see the available bandwidth as all remaining bandwidth, whereas for the new calls it is equal to available bandwidth minus the bandwidth occupied by suspended handoff calls. The state diagram of the requests is shown in fig 3.

Algorithm for Scheduler:

```
suspendlist: set of suspended HO connections
scheduledlist: set of scheduled new or HO connections
SetHotime=R(Th(p))
Set ReservationStartTime= Tr(p)
Do Every Frame
Set counter = 0
While (counter<MaxHORequest)
     Read Handover Request R(B,D,Hotime)
     Set R(Hotime)=R(Hotime)-1
     If R(Hotime) <0
       Delete R
     Else If R(Hotime) <=Reservation Start Time
       ProcessHORequest (R)
     End If
   counter=counter+1
 End While
Set Counter=0
While (counter<MaxNewRequest)
   Read New Request R(B,D)
   ProcessNewRequest(R)
   counter=counter+1
 End While
    Serve Requests
End Do
 Procedure ProcessHORequest(R)
   If R is not yet suspended
         Suspend (R)  'place in suspend list
         HOAvailBw =AvailBW
         NewAvailBW=AvailBW -Sum (suspendlist(bw))
    End If

    If HoAvailBw <= R(B)
```

```
    If R meets QoS Criteria (7)
        Schedule (R)    'place in scheduled list
      Else
        'meets BW but not QoS so dispose off
        Delete R
      End If
    Else
        ' does not meet bw requirement but may meet in
        'next frames ,so continue suspension
        Do Nothing
    End If
  End ProcessHORequest
Procedure ProcessNewRequest (R)
  If R meets QoS Criteria (7)
    Schedule(R)
  Else
    Delete R
  End If
End ProcessNewRequest
```

The state diagram of the requests is shown in fig 3.

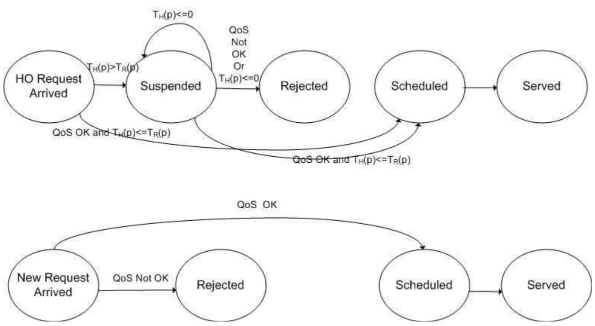

그림 3. Handover와 새로운 수락에 대한 상태천이도
Fig. 3. State diagram for handover and ner calls

Ⅵ. SIMULATION RESULTS

If an AMS expects its bandwidth request Brm to be admitted into the network then the best, worst and average case maximum delay Dm that AMS m can request successfully (without considering polling and channel access delays) is given by [Brm/Rm], [(Btot-Brm)/Rav+Brm/Rm] and [((Btot-Brm)/Rav+2Brm/Rm)/2] respectively expressed in the multiple of frame duration (In other words, $D_m \in n^* T_f$). Fig. 4 shows that the maximum delay for the admitted handover and

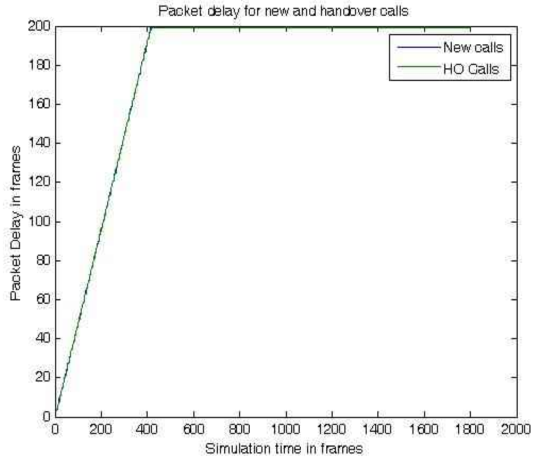

그림 4. 새로운 수락과 handover 수락에 대한 최대 패킷 지연 시간
Fig. 4. Maximum packet delay for new and handover calls

new calls are almost the same, since the handover-bias is done prior to scheduling, and not once the call is scheduled for transmission.

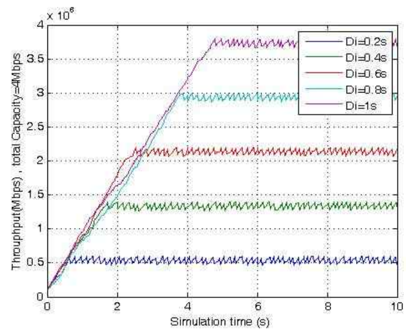

그림 5. 사용자 간 지연을 위한 처리량 지연 균형
Fig. 5. Throughput delay tradeoff different user defined delays

Fig 5 is the simulated result of 4Mbps system capacity for various user defined delays. It shows the delay throughput tradeoff. In general , the higher the user defined delay , the higher the throughput. The unused bandwidth can however be used by lower priority traffic such as best effort traffic.

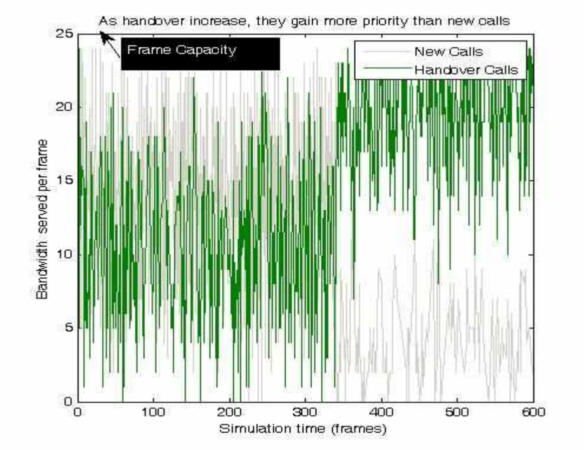

그림 6. Handover 수락 증가에 따른 새로운 수락의 우선 순위 증가

Fig. 6. As handover calls increase, they get higher priority than new

Fig 6 shows the handover-bias of the admission control. In the former frames the handover calls are requesting lower bandwidth. Later as the size of handover request grows, there is a bias in the system that rejects the new calls to provide more resource for the handover calls.

Ⅶ. CONCLUSIONS

In this paper we developed predictive reservation scheme for admitting real-time handover traffic into IEEE 802.16m based fourth generation IMT- Advanced Network, and guarantee them with QoS in terms of both minimum bandwidth and maximum tolerated delay. The admission control uses as its input the predicted handover time from the location and trajectory information. It then performs suspension scheduling with periodic reservation attempts for the handover calls. Whenever there are handover items in the suspension list, they are given more priority than the new calls. The algorithm however does not affect the minimum sustained rate of the new calls, as they must be met as signed in the SLA.

참 고 문 헌

[1] "IEEE 802.16m System Description Document (SDD)", *Website, IEEE*, 2009.

[2] "Requirements related to technical performance for IMT-Advanced radio interface(s)", *ITU*, 2008.

[3] "Guidelines for evaluation of radio interface technologies for IMT-Advanced", *ITU*, 2008.

[4] I Joe and M Shin, "A Mobility-based Prediction Algorithm with Dynamic LGD Triggering for Vertical Handover", *IEEE CCNC,* 2010 .

[5] "Draft Standard for Local and Metropolitan Area Networks: Media Independent Handover Services", *IEEE Computer Society*, June 2008.

[6] J.M. Bauset, J M G Guzman and V Pla, "Optimal Admission Control in Multimedia Mobile Networks with handover prediction", *Wireless Communications, IEEE, Vol.15* , Issue 5, 2008.

[7] K. Miyashita, T. Terada, and S. Nishio, "A Map Matching Algorithm for Car Navigation Systems that Predict User Destination", *22nd International Conference on Advanced Information Networking and Applications.*

[8] H. Ge X. Wen, W. Zheng, Z. Lu, B Wang ,"A History-Based Handover Prediction for LTE Systems", *International Symposium on Computer Network and Multimedia Technology*, 2009.

[9] 문재경, 오길남, 박재홍, 하영호, 김수중, "지상파 방송 채널에서 OFDM 시스템의 시간 평균 채널 추정기의 성능," *한국전자파학회논문지, 제10권, 제1호*, pp. 44-53, 1999. 2.

[10] K. Wongthavarawat and A. Ganz, "Packet Scheduling for QoS support in IEEE 802.16 broadband wireless access systems", *International Journal of Communication Systems, Vol. 16*, pp. 81-96,Feb 2003.

[11] D. Niyato and E. Hossain, "Connection Admission Control Algorithsm for OFDM Wireless Networks, *IEEE GLOBECOM, MO, USA,* 28 Nov-2 Dec 2005.

[12] J. Y. Lee, K. B. Kim, "Statistical Connection Admission Control for Mobile Wimax Systems",

[13] R. Iyengar, K. Kar and B. Sikdar, "Scheduling Algorithms for Point-to-Multipoint Operation in

IEEE 802.16 Networks", *Proceedings of 2nd Workshop on Resource Allocation in Wireless Networks (RAWNET), April* 2006.

[14] R. Iyengar, V. Sharma, K. Kar and B. Sikdar, "Analysis of Contention based Multi-Channel Wireless MAC for Point to multipoint Networks", *Proceedings of IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks, Buffalo,* June 2006.

[15] R. Iyengar, P. Iyer and B. Sikdar, "Delay Analysis of 802.16 based Last Mile Wireless Networks", *Proceedings of IEEE Globecom*, 2005.

[16] B.J. Chang, C. M. Chou and Y. H. Liang, "Markov Chain Analysis of Uplink subframe in polling-based WIMAX networks", *Computer Communications, Vol. 31, Issue 10,* pp. 2381-2390, June 2008.

Neeraj Poudyal

2002년 6월 : Bachelor of Computer Applications, MCRP University (공학사)
2005년 8월 : Masters of Computer Applications, Sikkim Manipal University of Health, Medical and Technological Sciences (공학석사)
2005년 9월~현재 : 한국항공대학교 정보통신공학과 박사과정
관심분야 : Admission Control, IMT-Advanced, IEEE 802.16m Wireless Communications and Networking Conference, 2008.

이 병 섭 (李 炳 燮)

1979년 2월 : 한국항공대학교 전자공학과 (공학사)
1981년 2월 : 서울대학교 전자공학과 (공학석사)
1992년 2월 : New Jersey Institude of Technology (공학박사)
1992년~현재 : 한국항공대학교 정보통신공학과 교수
관심분야 : 디지털 신호 처리, 위성 통신 , Adaptive Array